

Emergence, autopoïèse et autonomie

Jean-Pierre Müller, CIRAD-GREEN, Montpellier, France

Introduction

L'objectif de ce texte est de discuter la notion d'autonomie au regard des notions d'émergence et d'autopoïèse. En effet, la notion d'autonomie est beaucoup utilisée, notamment dans le domaine des systèmes multi-agents (Ferber, 1995), sans que le sens de ce terme, et surtout sa mise en oeuvre, ne soit vraiment débattue. Au sens strict, le mot autonome vient de "autos": soi-même et de nomos: "loi". Est donc autonome tout système qui se gouverne par ses propres lois. Si cette définition a un sens politique, pour un système comme un agent, un robot ou autre artefact, la signification est très loin d'être claire et comporte pêle-mêle:

- l'autonomie énergétique (usage de batteries),
- l'absence d'une intervention humaine (pas de pilotage à distance),
- le fait d'avoir des objectifs propres,
- etc.

donc, a priori, pas grand chose qui soit de façon évidente en lien avec la signification première de l'autonomie. Ceci n'empêche pas les tenants du domaine de prétendre que ce qui est fondamental dans les systèmes multi-agents, c'est l'autonomie des agents (en séparant les syllabes et en faisant de grand yeux!).

Sans vouloir apporter une réponse définitive à ce que peut vouloir dire "être autonome", nous allons à travers les notions d'émergence et d'autopoïèse, proposer un cadre d'analyse permettant de recouvrir divers sens d'autonomie et d'en discuter les implications, notamment en termes de réalisabilité. Cette dernière dimension est le propre de ce que j'appelle (avec (Simon 1969) mais pas dans le même sens!) les sciences de l'artificiel, à savoir les sciences, comme l'intelligence artificielle et la vie artificielle, qui prétendent aborder leur objet à travers la réalisation d'artefacts comme nouvel exercice de pensée. Cette discussion pourrait donc servir pour tracer les contours de l'autonomie artificielle.

Emergence

Pour l'émergence, nous allons reprendre la définition de (Müller, 2003) aussi citée dans (Dessales & al. 2007), à savoir qu'il y a émergence si on a:

- Un ensemble d'entités interagissantes dont les états et la dynamique est exprimée dans une théorie D;
- La production d'un phénomène (processus, état stable, invariant) qui est global relativement à l'ensemble précédent (ou à un de ses sous-ensembles);
- L'observation de ce phénomène via un mécanisme d'inscription, et son interprétation dans une autre théorie D';
- l'irréductibilité de D' à D (D' n'est pas une simple aggrégation/composition des éléments de la théorie D).

Un certain nombre de remarques s'imposent:

- l'existence d'un (ou plusieurs) observateur qui va respectivement définir les entités à considérer, observer l'existence d'un invariant et le théoriser de façon indépendante des descriptions des composants;
- l'existence de deux niveaux d'observation: le niveau micro des interactions entre entités et le niveau macro où s'observe le phénomène global;
- la mise en avant que l'observation passe par une matérialisation (inscription) sous une forme ou sous une autre du phénomène étudié, que ce soit au niveau micro comme au niveau macro. Cette matérialisation peut être sous la forme d'un contour matériel (par exemple une paroi cellulaire) ou nécessiter un système d'enregistrement plus ou moins sophistiqué (par exemple, l'enregistrement de la valeur de variables au fil du temps, voire son dessin sur un support, pour s'apercevoir qu'il y a une régularité).

A noter qu'on ne parle pas de système ni, a fortiori, de système complexe car le phénomène émergent peut se produire au sein de l'ensemble des entités interagissantes (comme les "glisseurs" dans le jeu de la vie) ou définir l'ensemble des entités dont l'interaction nous intéresse (comme la cellule relativement aux processus biochimiques qu'elle met en oeuvre). Cependant, dans un cas comme dans l'autre, une certaine identité de l'ensemble ou d'un sous-ensemble des entités considérées, caractérisée par une ou des invariances spatio-temporelles, apparaît aux yeux de l'observateur.

On distingue différentes formes d'émergence parmi lesquelles l'émergence faible dans laquelle le phénomène émergent n'existe que dans l'oeil de l'observateur extérieur, et l'émergence forte dans laquelle le phénomène global rétroagit sur les composants de l'ensemble observé. Ce concept de rétroaction du niveau macro sur le niveau micro doit être finement exploré. Dans l'exemple des fourmis qui, dans le cheminement entre le nid et la source de nourriture, déposent des phéromones qui vont finir par tracer un chemin, il y a bien rétroaction du chemin produit sur les fourmis puisque cette trace de phéromone va recruter de nouvelles fourmis pour le transport. Mais cette trace n'est perçue que localement par les fourmis. Il n'y a pas de raison de supposer que chaque fourmi perçoive le chemin dans son ensemble, ce qui serait le cas si elles en faisaient une carte, au moins dans leur mémoire. Donc si, dans cet exemple, le phénomène global rétroagit bien sur le phénomène local, il ne rétroagit pas "en tant que" phénomène global. Si c'était le cas, les fourmis pourraient s'en servir pour anticiper un autre chemin avant même de l'avoir parcouru (ce que les navigateurs portugais ont pu faire avec le succès que l'on connaît (Hutchins, 1995)).

A partir de cette définition, nous aurions déjà un premier niveau d'autonomie qui est la constitution spontanée d'une forme d'identité (une forme de loi très simple qui consiste à rester égal à soi-même, d'exprimer un invariant). Maturana (Maturana 1988) interroge d'ailleurs cette notion d'identité en distinguant organisation et structure:

- une organisation est un ensemble de rôles et de relations qui définissent une identité (par exemple, être une table comme relation entre un rôle de support et un rôle de stabilité)
- une structure est une réalisation concrète d'une organisation (par exemple, une pierre plate ou un plateau avec trois pieds)

A noter qu'une organisation peut potentiellement spécifier une infinité de structures possibles. Tant qu'un objet dans ses transformations successives maintient son

organisation, il conserve son identité au sens de Maturana. On remarquera que cette définition introduit deux formes d'identité:

- une identité matérielle: le système maintient ses composants où les renouvelle dans une continuité spatio-temporelle
- une identité de type: son organisation se maintient également et l'identifie (on peut le dire!) à un type (par exemple, une table)

A noter que l'organisation d'un phénomène émergent (ou pas, comme dans le cas de la table) n'est qu'une condition de possibilité pour être identifié matériellement ou typologiquement par un observateur extérieur. Un observateur extérieur reste "libre" de considérer un objet combinant un rôle de support et de stabilité comme une table ou non.

Nous retiendrons pour la suite de l'exposé, les notions d'émergence faible et forte et l'esquisse d'une forme d'identité.

Autopoïèse

Un système autopoïétique est défini de la façon suivante (Maturana 1988, Maturana & Varela 1994):

- Une machine autopoïétique est un réseau de processus de production (création, transformation et destruction) de composants qui produisent les composants qui:
 - 1) à travers leurs interactions et transformations régénèrent continuellement et réalisent le réseau de processus (relations) qui les a produit;
 - 2) constitue le réseau comme une unité concrète dans l'espace dans laquelle les composants existent en spécifiant le domaine topologique de sa réalisation en tant que réseau.

Le modèle minimal d'un système autopoïétique est une cellule dont la membrane est composé d'un constituant C se dégradant en D et baignant dans un milieu riche en molécules A. Ces molécules A peuvent franchir la membrane et être transformées au sein de la cellule en molécules B pour lesquelles la membrane est imperméable. B peut s'intégrer à la membrane pour se transformer en C. Si le débit d'entrée de A et sa conversion en B sont suffisamment grands devant le coefficient de dégradation de C en D alors la cellule se maintient au cours du temps. Ce modèle montre l'importance de la frontière du système (ici la membrane). Si celle-ci disparaît le métabolisme et le système entier s'effondre. Il s'agit d'un cercle vicieux : si B s'échappe, sa concentration diminue de tel sorte que la membrane se dégrade de plus en plus vite et que la perte en B augmente. Le métabolisme et la membrane dépendent l'un de l'autre. La structure ne peut se maintenir sans le flux (Bourgine & Stewart, 2004).

Si nous analysons cette définition au regard de la définition d'émergence, nous avons:

- Un ensemble de composants interagissants dont les états et la dynamique (ici de création, transformation et destruction) est exprimée dans une théorie D. Dans l'exemple:
 - les composants sont les molécules A, B, C, D;
 - les interactions sont les transformations $A \rightarrow B$, $B \rightarrow C$ et $C \rightarrow D$ plus le fait que C repousse B, aggrège avec C et est indifférent à A (et probablement D);

- la production d'une topologie sous la forme d'une co-production d'une paroi et d'un différentiel de concentration de certaines substances entre un intérieur et un extérieur;
- la théorisation de cette topologie sous la forme (par exemple) d'une activité métabolique de maintenance d'une identité "cellulaire" dans le temps;
- l'irréductibilité de cette auto-maintenance à une simple composition des interactions citées, ne fût-ce que par la création d'une forme spatiale.

Il y a bien une rétroaction du niveau macro vers le niveau micro par la création d'une paroi qui va avoir comme effet de maintenir un différentiel de composition entre un intérieur et un extérieur mais il ne s'agit bien évidemment pas d'une émergence forte.

Nous avons donc deux niveaux:

- le niveau micro de processus analogues à des transformations chimiques et des interactions physiques;
- le niveau macro d'une organisation topologique de processus d'absorption des molécules A, de rejet (?) des molécules D et de métabolisme interne ($A \rightarrow B \rightarrow C \rightarrow D$).

Comme le souligne Varela, au niveau micro, ce qui se passe entre la "cellule" et l'extérieur peuvent être considérés comme des perturbations alors que le niveau macro introduit un niveau interprétable en termes de l'organisation émergente. Par exemple, on pourrait dire que la cellule a "trop à manger", "pas assez à manger", s'adapte à des variations du milieu, etc.

Nous obtenons un autre niveau d'autonomie par deux dimensions:

- un renforcement de la notion d'identité, puisque cette identité n'est pas seulement constatée mais activement maintenue par un réseau de processus qui définissent un "domaine topologique de réalisation" comme le dit la définition d'autopoïèse;
- les conditions de possibilité d'interpréter les interactions aux niveaux micro comme ayant un "sens" pour cette organisation (a minima, favorable, défavorable, voire "létal").

On est très proche de la définition d'autonomie puisque, d'une certaine façon, le système se dote bien de ses propres lois, à savoir son organisation, et, qu'en plus, ces lois définissent son interprétation des interactions. Cependant, c'est l'observateur extérieur qui va expliciter ces lois et construire l'interprétation des interactions en leurs termes et pas le système lui-même. Et c'est encore moins le système qui va pouvoir éventuellement les choisir. C'est le dépassement de ces limites, et donc la possibilité de construire une image de soi et, partant de choisir, que nous allons esquisser maintenant.

Autonomie

De ce qui précède, nous avons donc à esquisser (nous n'en ferons pas davantage) les deux caractéristiques qui manquent, et à notre sens sont liées, pour obtenir ce que l'on pourrait appeler une véritable autonomie:

- que l'organisation deviennent des lois pour le système lui-même, et pas seulement pour un observateur extérieur;
- que le système puisse changer ces lois, ou les choisir.

La première condition fait écho à la définition de l'émergence forte. En effet, il y a émergence forte si le phénomène global est perceptible par les éléments qui le compose. Cela reviendrait à pouvoir constituer au sein même du système émergent une représentation de cette émergence ce qui pose deux questions:

- la question de la matérialité de cette représentation puisque nous avons vu qu'une émergence passe nécessairement par une inscription, une réification au sens propre, pour pouvoir être identifiée en tant que telle;
- la question plus subtile du processus d'internalisation du regard extérieur pour deux raisons:
 - l'inscription ne suffit pas, elle doit aussi être interprétée dans une théorie D' par un observateur extérieur;
 - et donc se pose la question de comment le système peut se décrire comme si il était à l'extérieur de lui-même.

La première question est partiellement répondue par la nécessaire inscription matérielle du phénomène émergent comme condition de possibilité pour son interprétation comme un tout. Par exemple, on pourrait se demander dans quel sens le code génétique constituerait une forme d'inscription du fonctionnement de la cellule et de l'organisme (Leroi-Gourhan, 1964-65).

En ce qui concerne l'interprétation de cette inscription, il est clair que sa nature même d'inscription lui donne le même statut que n'importe quelle source d'interactions au niveau micro. Ces interactions acquerront un sens que par rapport à une organisation qui leur confère la possibilité d'être interprétées fonctionnellement. Il reste à imaginer comment donner à cette interprétation un statut d'auto-référence, donc d'interroger, en général, comment se construit une auto-référence, une sorte de "conscience de soi". C'est le genre de questions qu'aborde la littérature notamment en phénoménologie et ceci depuis Hegel (Hegel 1807) mais que nous n'avons pas eu le temps d'aborder en détail.

Il nous semble néanmoins que cette auto-référence esquissée ici est une condition nécessaire, sinon suffisante, pour introduire la notion de choix. De même que la carte du chemin de la fourmi lui donne les conditions de possibilité de prendre un autre chemin, qui plus est jamais parcouru au préalable, la conscience des lois qui nous régissent individuellement et collectivement nous donne les conditions de possibilité de leur violation, de leur changement et de leur choix, constituant ainsi une véritable autonomie.

Conclusion

En partant de la définition d'autonomie, nous avons proposé des niveaux successifs de sophistication des systèmes qui nous paraissent successivement nécessaires pour pouvoir saisir la notion d'autonomie au sens plein du terme. De cette discussion, il en ressort que les notions d'émergence (et donc de système complexe), puis d'autopoïèse comme un cas particulier d'émergence, puis d'émergence forte comme préalable à une forme de conscience semblent, conceptuellement, les étapes obligées pour en définir les conditions de possibilités, et donc, partant de là, de réalisabilité.

Force est de constater qu'un agent logiciel ne se constitue pas de façon émergente à partir d'un substrat, est encore moins autopoïétique, ni ne fait émerger une forme

d'auto-référentialité. Si on considère que ce sont des passages obligés, la propriété d'autonomie attribuée aux agents est donc au mieux métaphorique. Mais on sait toute la richesse que peut amener la métaphore poussée jusqu'à ses retranchements. C'est ce que nous avons essayé de proposer ici. L'émergence instaurant une forme d'indépendance théorique d'un niveau par rapport à l'autre, il est peut-être aussi possible d'explorer l'auto-référentialité sans avoir à construire un système autopoïétique. Il resterait à discuter si l'auto-référentialité telle qu'esquissée ici peut ou non se réduire à la réflexivité dans les systèmes informatiques (Maes & Nardi, 1988).

Références

- Paul Bourguin and John Stewart, Autopoiesis and Cognition. *Artificial Life* 10 327-345, 2004.
- Jean Louis Dessalles, Jean Pierre Müller and Denis Phan, Emergence in Multi-agent Systems: Conceptual and Methodological Issues, in *Agent-based Modelling and Simulation in the social and human sciences*, Bardwell Press, 2007
- Jacques Ferber, *Les systèmes multi-agents*, InterEditions, 1995
- Georg Wilhelm Friedrich Hegel, *La Phénoménologie de l'esprit*, Bamberg und Würzburg, 1807
- Edwin Hutchins, *Cognition in the wild*, MIT Press, 1995
- André Leroi-Gourhan, *Le geste et la parole* (1 et 2), Albin-Michel, 1964-1965
- Pattie Maes and Daniele Nardi, *Meta-level architectures and reflection*, North-Holland, 1988
- Humberto Maturana, The Ontology of observing, in *Text in Cybernetics*, American Society for Cybernetics, 1988
- Humberto Maturana, Francisco J. Varela, *L'arbre de la connaissance, racines biologiques de la compréhension humaine*, Addison-Wesley 1994
- Jean-Pierre Müller, Emergence of Collective Behaviour and Problem Solving, in *ESAW* 2003, LNAI 3071, Springer, pp. 1-20
- Herbert Simon, *The science of the artificial*, MIT Press, 1969